



Statistik Deskriptif

Pada bab terdahulu telah dibahas penggunaan fasilitas OLAP dari SPSS. Dengan OLAP, sejumlah variabel dengan mudah dapat dikaitkan untuk memperoleh sejumlah informasi. Untuk penyajian informasi yang tidak kompleks, OLAP sudah memadai pada banyak kegiatan pengolahan data, seperti penyajian komposisi penjualan berdasar wilayah, produksi beras berdasar daerah dan tingkat curah hujan, motivasi karyawan berdasar pengalaman kerja dan gaji yang diperoleh, dan lain sebagainya.

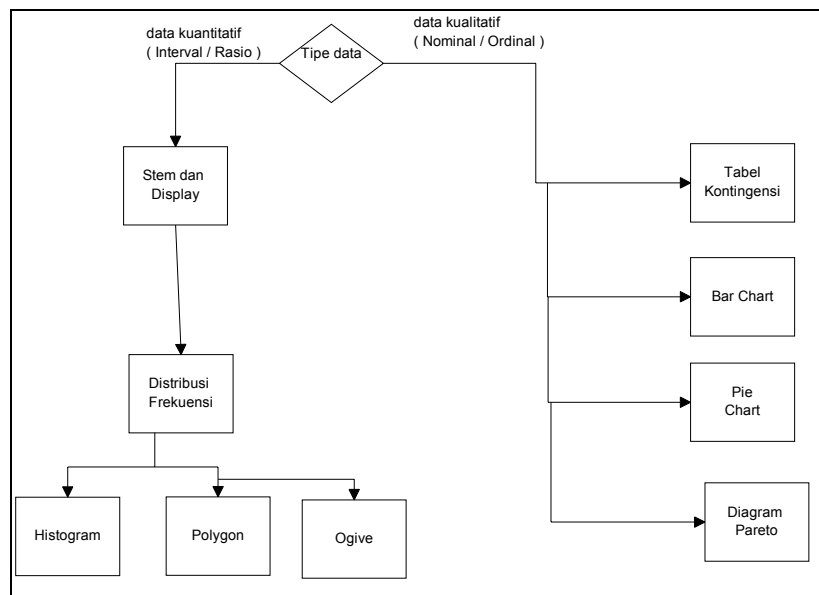
Namun demikian, dalam banyak pengolahan data dibutuhkan pula informasi yang sifatnya lebih rinci dan mendalam; atau diinginkan tampilan informasi yang sifatnya khusus. Misalkan akan ditampilkan ringkasan data untuk produksi padi di daerah X saja selama satu bulan terakhir; atau ringkasan total pendapatan para salesman dalam daerah pemasaran tertentu selama satu minggu. Dapat pula informasi yang diminta adalah apakah data tertentu berdistribusi normal ataukah tidak, dengan tujuan kegunaan kegiatan inferensi tertentu. Pada kasus-kasus semacam ini, fasilitas OLAP kurang memadai, sehingga diperlukan fasilitas-fasilitas tambahan atau yang bersifat khusus. Namun semua fasilitas tambahan tersebut tetap dalam konteks mendeskripsikan data yang telah dikumpulkan, dan tidak melakukan inferensi apa pun terhadap data yang ada, seperti misalkan meramalkan data di masa mendatang atau pengujian dua kelompok data atau lebih. Itulah ciri utama **statistik deskriptif**, yakni mendeskripsikan atau menjelaskan atau menggambarkan data tanpa menarik kesimpulan apa pun.

Hal itu dapat diumpamakan seseorang yang melihat sebuah panorama alam, kemudian menjelaskan bahwa dia melihat sebuah gunung, jalan berkelok-kelok di bawahnya, ada sejumlah rumah, sejumlah besar pohon, dan item-item yang lain. Orang tersebut sekadar menggambarkan kemudian dapat menambah dengan melaporkan panjang sungai diperkirakan sekian kilometer, ada seratus rumah besar dan sepuluh rumah kecil, ada lebih dari seribu pohon cemara, dan seterusnya. Namun ia tidak melontarkan pendapat atau

menarik kesimpulan apa pun; sebagai contoh, ia tidak mengatakan sebaiknya rumah jangan ada di lereng gunung karena alasan keselamatan, ia juga tidak melontarkan dugaan bahwa semakin banyak rumah dibangun akan semakin banyak pohon di sekitar lereng gunung akan ditebang, dan kesimpulan lainnya.

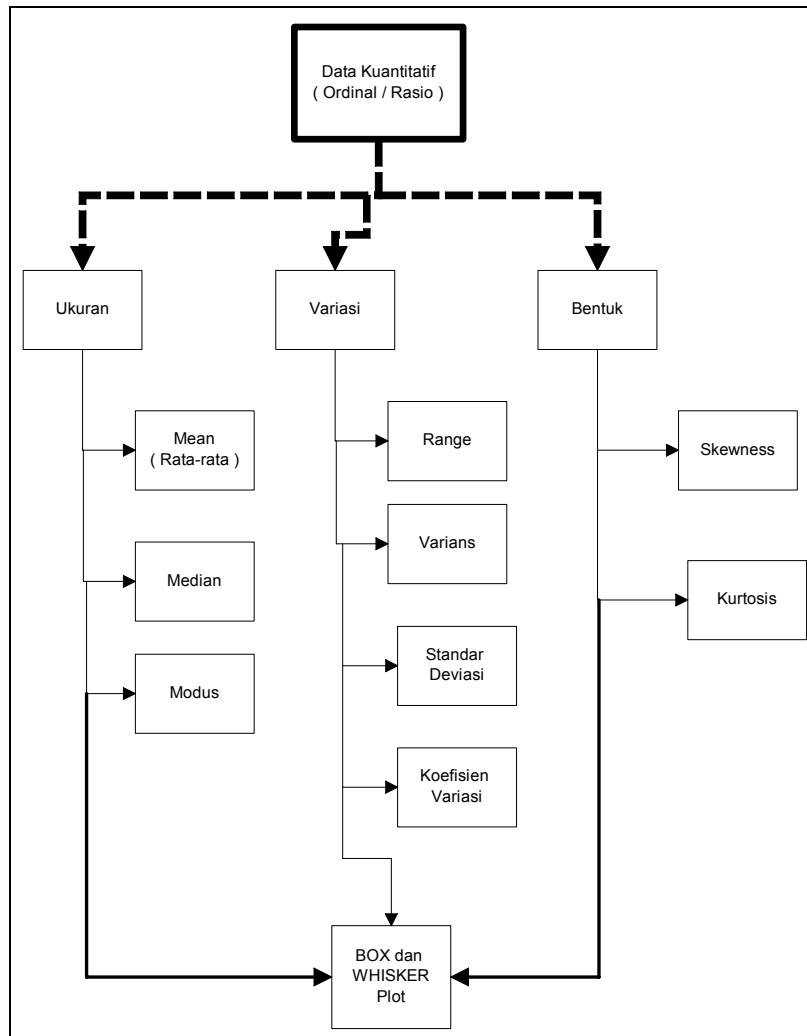
Secara umum, kegiatan dalam statistik deskriptif dapat dibagi dalam dua bagian dasar berikut ini:

- **Menyajikan Data;** data bisa disajikan dalam bentuk TABEL dan GRAFIK.



Penyajian grafik dalam bentuk bar, pie, line dan sebagainya, bisa dilihat pada bab yang lain.

- **Meringkas dan menjelaskan Data;** data kuantitatif bisa diringkas dan disajikan dalam 3 hal utama:
 1. Letak data (disebut ‘Central Tendency’).
 2. Variasi data.
 3. Bentuk data.



SPSS menyediakan sejumlah fasilitas khusus untuk kegiatan mendeskripsikan data tanpa melakukan inferensi pada data. Fasilitas-fasilitas yang sifatnya lebih rinci atau untuk kegunaan khusus pada statistik deskriptif di software SPSS adalah:

- SUMMARIES, yang berisi ringkasan-ringkasan penting data, yang dapat dibagi menjadi:
 - ringkasan per kasus (case summaries)
 - ringkasan berbasis baris (report summaries in row)

Laporan (report) ini menampilkan laporan dengan berbagai ringkasan statistik terletak di baris.

- ringkasan berbasis kolom (report summaries in column)

Laporan (report) ini menampilkan laporan dengan berbagai ringkasan statistik terletak di kolom.

➤ Menu DESCRIPTIVE STATISTICS, yang dapat dirinci lagi menjadi:

- FREQUENCIES

Tampilan berbagai ringkasan statistik yang disertai dengan grafik-grafik statistik sederhana.

- DESCRIPTIVES

Fasilitas ini dapat menampilkan berbagai ringkasan statistik beberapa variabel pada satu tabel, dengan disertai nilai data yang terstandarisasi (nilai z).

- EXPLORE

Fasilitas ini menampilkan ringkasan statistik dan grafik statistik yang cukup kompleks, dengan kegunaan yang paling luas untuk kegiatan statistik deskriptif, seperti menguji normalitas data, identifikasi ada tidaknya data yang ekstrem dan outlier, dan perbedaan statistik antar-kelompok data.

Dari penjelasan di atas, terlihat bahwa penggunaan beberapa kelompok fasilitas tersebut pada dasarnya tergantung tujuan pengolahan data. Fasilitas SUMMARIES tentu yang paling sederhana, sedangkan kelompok DESCRIPTIVES STATISTICS untuk pengolahan lanjutan, dengan fasilitas EXPLORE adalah yang paling kompleks. Namun demikian, semua fasilitas tersebut dalam praktik akan saling melengkapi, dan pada umumnya dilakukan laporan gabungan untuk pengolahan file yang sama. Sebagai contoh, pengolahan data penjualan per daerah dapat ditampilkan ringkasannya dengan menu SUMMARIES, yang kemudian dirinci lagi untuk pengujian adanya data ekstrem dengan menu EXPLORE, dan sejumlah grafik data penjualan dapat ditampilkan lewat menu FREQUENCIES.

Berikut akan dibahas berbagai kasus untuk penggunaan fasilitas-fasilitas tersebut.

2.1 CASE SUMMARIES

Fasilitas ini bertujuan menyajikan ringkasan suatu variabel, dengan tampilan setiap kasus. Jika suatu variabel mempunyai 100 kasus, maka Case Summaries akan menampilkan setiap kasus tersebut, sehingga akan ada 100 baris pada laporan, dengan pembagiannya sesuai kriteria yang ditentukan.

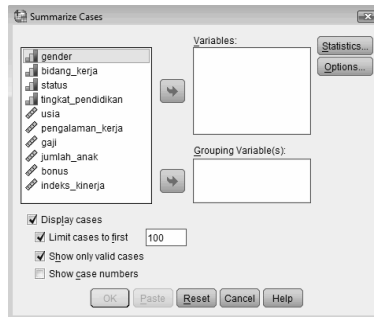
Data yang ditampilkan pada CASE SUMMARIES bisa berupa data kualitatif atau data kuantitatif.

KASUS 1

Dari data karyawan PT MAKMUR RASA akan dibuat Case Summaries dari variabel STATUS berdasar Bidang Kerja.

Langkah:

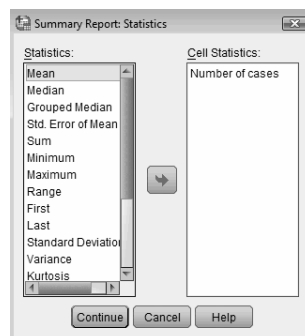
- Buka file **Karyawan**.
- Menu **Analyze → Reports → Case Summaries....** Tampak di layar:



Gambar 2.1 Kotak dialog SUMMARIZE CASE

Pengisian:

- **Variable(s)** atau tampilan pada kolom. Pilih variabel **Status**.
- **Grouping Variable(s)** atau tampilan pada baris. Pilih variabel **bidang_kerja**.
- Klik mouse pada pilihan STATISTICS, maka tampak di layar:



Gambar 2.2 Kotak dialog STATISTICS

Pilihan Statistics meliputi berbagai ukuran untuk menggambarkan data (statistik deskriptif). Karena kedua variabel yang dimasukkan adalah data kualitatif, maka di sini hanya dipilih *default* yang ada, yaitu **Number of Cases** (jumlah kasus).

Tekan CONTINUE untuk kembali ke kotak dialog utama.

Abaikan bagian yang lain, dan tekan OK untuk proses data.

Output

Summarize

Case Processing Summary ^a						
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Status Karyawan * Bidang	75	100,0%	0	,0%	75	100,0%

a. Limited to first 100 cases.

Output menunjukkan semua kasus (75 baris data karyawan) telah diproses semuanya. Selanjutnya akan ditampilkan komposisi dari ke 75 karyawan tersebut berdasar bidang kerjanya. Sebagai contoh, berikut output ringkasan untuk bidang kerja administrasi.

Case Summaries ^a				Status Karyawan
Bidang	Administrasi	1		Belum Menikah
		2		Belum Menikah
		3		Belum Menikah
		4		Belum Menikah
		5		Belum Menikah
		6		Belum Menikah
		7		Menikah
		8		Menikah
		9		Menikah
		10		Menikah
		11		Menikah
		12		Belum Menikah
		13		Belum Menikah
		14		Menikah
		15		Menikah
		Total	N	15

Dari output terlihat ada 15 karyawan di bidang kerja administrasi; karena tampilan yang dibutuhkan hanya mengenai status karyawan, terlihat karyawan bidang administrasi dengan nomor urut 1 berstatus belum menikah, sedangkan karyawan dengan nomor urut 7 berstatus telah menikah. Total ada 8 karyawan yang belum menikah dan 7 karyawan telah menikah.

Dari kasus sederhana di atas terlihat kegunaan CASE SUMMARIES adalah meringkas isi variabel untuk kelompok (grup) tertentu dengan memerhatikan isian variabel grupnya. Dalam kasus di atas, CASE SUMMARIES meringkas isi variabel bidang kerja; karena ada lima bidang kerja, maka akan ada lima *summaries*, dengan tampilan inti adalah status tiap kasus (karyawan), karena *grouping variable* adalah status.

NB: simpan file dengan nama **CASE SUMMARIES 1**.

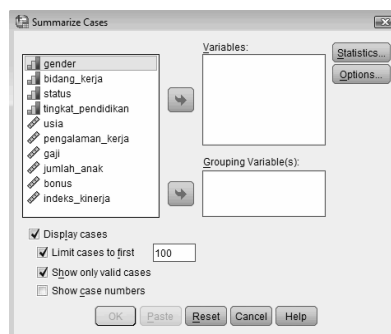
KASUS 2

Dari data karyawan PT MAKMUR RASA yang memuat kasus dengan *missing data* (data tidak didapat) akan dibuat Case Summaries dari variabel GENDER karyawan dan STATUS karyawan berdasar usia, pengalaman kerja, dan indeks kinerja mereka selama ini.

Perhatikan! Pada kasus ini data agak berbeda, karena pada variabel indeks kinerja ada sejumlah data kinerja karyawan yang tidak tersedia (disebut dengan *missing data*). Karena itu, nama file yang digunakan dibedakan, yakni karyawan 2.

Langkah:

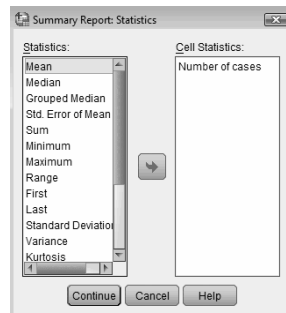
- Buka file **Karyawan 2**.
- Menu **Analyze → Reports → Case Summaries....** Tampak di layar:



Gambar 2.3 Kotak dialog SUMMARIZE CASE

Pengisian:

- **Variable(s)** atau tampilan pada kolom. Pilih variabel **usia**, **pengalaman_kerja** dan **indeks_kinerja**.
- **Grouping Variable(s)** atau tampilan pada baris. Pilih variabel **gender** dan **status**.
- Klik mouse pada pilihan STATISTICS, maka tampak di layar:



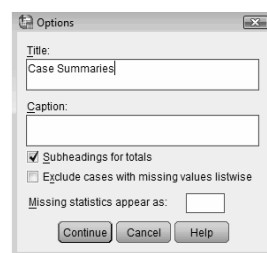
Gambar 2.4 Kotak dialog STATISTICS

Pengisian:

Karena ketiga variabel yang dimasukkan adalah data kuantitatif, maka di sini dapat dipilih sejumlah besaran statistik yang relevan, yang pada kasus ini adalah **mean** (rata-rata).

Tekan CONTINUE untuk kembali ke kotak dialog utama.

Kemudian tekan OPTIONS, sehingga tampak di layar kotak dialog:



Gambar 2.5 Kotak dialog OPTIONS

Pengisian:

Kotak OPTIONS adalah sejumlah pilihan untuk melengkapi tampilan CASE SUMMARIES namun tidak bersifat prinsip. Bagian pertama adalah judul output (TITLE), yang dapat diisi dengan kalimat-kalimat bebas yang menggambarkan hasil pengolahan data.

Sedangkan bagian CAPTION dapat diisi keterangan tambahan untuk melengkapi isian TITLE. Pada kasus ini kedua pilihan tersebut diabaikan saja.

Sekarang arahkan kursor pada isian MISSING STATISTICS APPEAR AS. Bagian ini berhubungan dengan isian tanda untuk data yang tidak ada nilainya. Untuk keseragaman, isi kotak di kanan bagian itu dengan tanda ***.

Tekan CONTINUE untuk kembali ke kotak dialog utama; kemudian tekan OK untuk proses data.

Output (bagian dua):

Case Summaries ^a							
					usia karyawan (tahun)	Pengalaman Kerja	Indeks kinerja karyawan dengan skala 0-10
Gender	Wanita	Status Karyawan	Belum Menikah	1	23	2	4,75
				2	21	2	6,00
				3	22	3	***
				4	24	5	8,08
				5	25	4	4,13
				6	25	4	***
				7	24	3	4,86
				8	20	2	5,12
				9	22	5	6,21
				10	24	4	6,15
				11	25	2	6,21
				12	26	5	8,29
				13	26	4	4,18
				14	30	5	7,92
				15	22	2	4,78
				16	22	6	8,70
				17	24	1	4,75
				18	25	2	2,05
				19	27	5	8,49
				20	27	5	6,52
				21	29	6	9,34
		Total	Mean	24,43	3,67	6,1332	

Pada output tampak empat kolom. Kolom pertama dilakukan pembagian antara gender karyawan untuk setiap status yang ada; karena ada dua gender (laki dan wanita) serta dua status karyawan (menikah dan belum menikah), maka akan terdapat empat kelompok ringkasan. Sedangkan tiga kolom berikutnya adalah data-data kuantitatif untuk setiap kasus (karyawan), yakni usia, pengalaman kerja, dan indeks kinerja mereka satu per satu.

Sebagai contoh, pada baris pertama adalah data karyawan pertama yang mempunyai gender wanita dan berstatus belum menikah. Usia karyawan

tersebut adalah 23 tahun, berpengalaman kerja sebanyak dua tahun, dan mempunyai indeks kinerja sebesar 4,75. Demikian seterusnya untuk data yang lainnya, sampai 75 data semua karyawan.

Berbeda dengan kasus sebelumnya, pada kasus ini ada sejumlah data yang tidak tersedia, khususnya pada variabel indeks kinerja. Pada kelompok karyawan wanita dan berstatus belum menikah, dari 21 karyawan yang ada, terdapat dua data, yakni karyawan nomor 3 dan nomor 6, yang tidak mempunyai data indeks kinerja. Perhatikan tanda *** untuk data missing, sesuai dengan tanda yang dimasukkan pada kotak dialog OPTIONS.

Pada baris terakhir untuk kelompok ini, terdapat keterangan MEAN untuk variabel usia adalah 24,43 yang berarti rata-rata usia karyawan wanita dan belum menikah adalah 24,43 tahun. Demikian pula untuk rata-rata pengalaman kerja mereka yang sebesar 3,67 tahun; sedangkan rata-rata indeks kinerja mereka adalah 6,1332, yang didapat dengan *tidak mengikutsertakan data missing*, atau dalam kasus ini hanya melakukan rerata untuk 19 data saja.

NB: simpan file dengan nama **CASE SUMMARIES 2**.

2.2 CASE SUMMARIES In Row

Fasilitas ini hampir sama dengan CASE SUMMARIES yang sudah dibahas di atas, yakni bertujuan menyajikan ringkasan suatu variabel, dengan tampilan disajikan per baris.

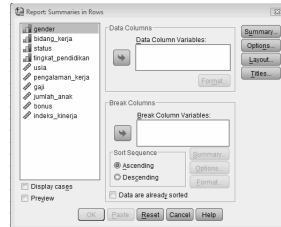
Data yang ditampilkan pada CASE SUMMARIES bisa berupa data kualitatif atau data kuantitatif.

KASUS

Dari data karyawan PT MAKMUR RASA akan dibuat Case Summaries In Row dari gaji dan usia karyawan berdasar status mereka.

Langkah:

- Buka file **Karyawan**.
- Menu **Analyze → Reports → Case Summaries in Row....** Tampak di layar:



Gambar 2.6 Kotak dialog SUMMARIZE CASE IN ROWS

Pengisian:

- **Data Column Variable(s)** atau tampilan pada kolom. Pilih variable **gaji** dan **usia**.
- **Break Column Variable(s)** atau pemisahan isian data. Pilih variabel **status**.

Aktifkan pilihan **Display Cases** yang ada di bagian kiri bawah kotak dialog. Dengan pilihan ini, output akan menampilkan nomor-nomor kasus yang sesuai dengan kriteria yang telah diinputkan ke dalam kotak dialog di atas.

Abaikan bagian yang lain, dan tekan OK untuk proses data.

Output (ditampilkan sebagian):

Report		
Page 1		
usia		
Status	Gaji per	karyawan
Karyawan	bulan	(tahun)
<hr/>		
Belum		
Menikah		
	1190000	23
	1020000	21
	1140000	22
	1350000	24
	1160000	25
	1200000	25
	1090000	21
	1125500	26
	1200000	20
	1455000	24
	930500	24
	960000	20
	1020000	21
	1371000	30
	1020000	20
	1260000	22

Terlihat output gaji dan usia untuk setiap karyawan ditampilkan per baris (row) dengan pembagian berdasar status mereka. Pada bagian pertama, pembagian berdasar karyawan dengan status belum menikah terlebih dahulu; pada baris pertama, ada karyawan yang belum menikah dengan gaji Rp. 1.190.000,- per bulan dan berusia 23 tahun. Demikian seterusnya untuk data yang lain, dengan memerhatikan status mereka.

Simpan file dengan nama **case summaries rows**.

2.3 CASE SUMMARIES In Column

Fasilitas ini sama dengan CASE SUMMARIES IN ROW, hanya sekarang ringkasan ditampilkan per kolom.

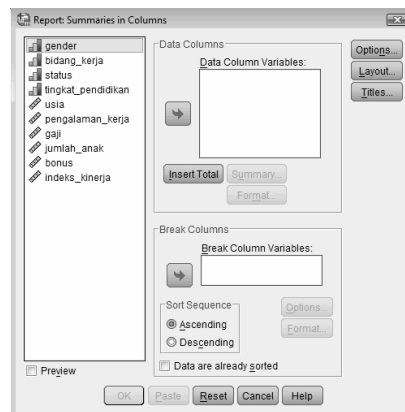
Data yang akan diproses bisa data kualitatif atau data kuantitatif.

KASUS

Dari data karyawan yang ada di PT MAKMUR RASA, akan dibuat ringkasan variabel gaji dan usia berdasar gender karyawan kemudian variabel gender dirinci lagi berdasar status karyawan.

Langkah:

- Buka file **Karyawan**.
 - Menu **Analyze → Reports → Case Summaries in Columns...**
- Tampak di layar:



Gambar 2.7 Kotak dialog SUMMARIZE CASE IN COLUMNS

Pengisian:

- **Data Columns Variable** atau tampilan variabel pada kolom. Pilih variabel **gaji** dan **usia**.
- Klik tombol **SUMMARY** (ada pada bagian Data Columns) dan kemudian pilih **Mean of Values** (untuk menghitung rata-rata gaji dan usia).

Tekan tombol **Continue** untuk melanjutkan.

Setelah pengaktifan summary Mean, otomatis variabel Gaji dan Usia menjadi gaji:mean dan usia:mean. Jika ini tidak diaktifkan, otomatis SPSS memilih default, yaitu Sum atau jumlah, yang tentunya tidak relevan memilih jumlah usia karyawan.

- **Break Columns** atau tampilan pada baris (yang membagi isi kolom). Pilih variabel **gender** dan **status**.
 - Klik tombol **FORMAT** (ada pada bagian Break Columns), dan kemudian aktifkan **Value Labels** pada bagian Column Content.
- Lalu tekan tombol **Continue** untuk melanjutkan.

Abaikan bagian lain dan tekan OK.

Output:

Report

Page 1			
usia			
Gaji per karyawan			
(tahun)			
Gender	Status Karyawan	Mean	Mean
Wanita	Belum Menikah	1214619	24
	Menikah	1408708	29
Pria	Belum Menikah	1192964	23
	Menikah	1401179	28

Simpan file dengan nama **CASE SUMMARIES COLUMN**.

Terlihat Karyawan Wanita yang menikah mempunyai rata-rata gaji yang paling besar, namun juga mempunyai rata-rata usia yang paling tinggi dibanding kelompok lainnya. Demikian bisa dilakukan analisis yang lain berdasar output di atas.

Dengan CASE SUMMARIES, baik itu ditampilkan per baris atau kolom, lebih menekankan pada cara penyajian informasi. Berikut akan dijelaskan fasilitas-fasilitas SPSS yang lebih mendetail pada besaran-besaran penting statistik.

2.4 FREQUENCIES

Fasilitas FREQUENCIES akan menggambarkan data dalam berbagai ukuran pusatnya, seperti Mean, Median, Persentil, dan lainnya.

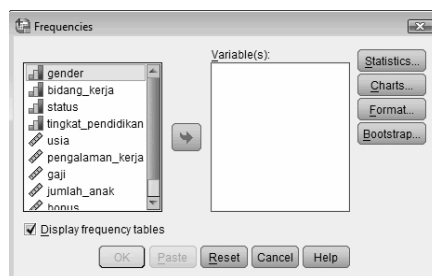
Data pada fasilitas ini bisa berupa data kualitatif atau data kuantitatif.

KASUS 1

Dari data karyawan PT MAKMUR RASA, manajer SDM ingin mengetahui lebih detail data usia karyawan. Untuk itu, akan dibuat Tabel Frekuensi serta melakukan deskripsi data pada variabel tersebut.

Langkah:

- Buka lembar kerja **Karyawan**.
- Menu **Analyze** → **Descriptive Statistics** → **Frequencies...**

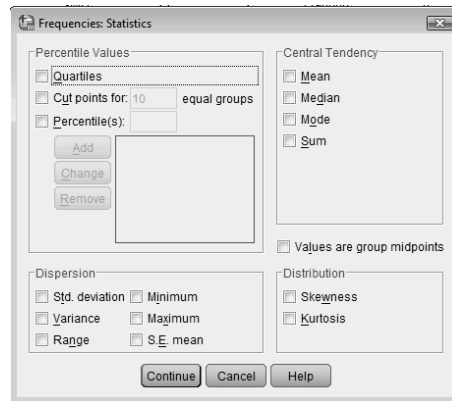


Gambar 2.8 Kotak dialog FREQUENCIES

Pengisian:

- **Variable(s)** atau variabel yang akan dimasukkan. Sesuai kasus, masukkan variabel **usia**.

- Buka pilihan STATISTICS, maka tampak di layar:



Gambar 2.9 Kotak dialog STATISTICS

Pengisian:

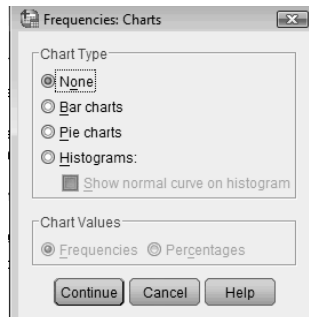
- **Percentiles Values** atau Nilai persentil, untuk keseragaman, klik mouse pada **Quartiles** dan **Percentile(s)**. Kemudian pada kotak di samping kanan Percentiles ketik **10**, lalu tekan **Add**. Sekali lagi ketik **90** pada kotak terdahulu, dan klik lagi tombol **Add**. Pengerjaan ini dimaksudkan untuk membuat nilai persentil pada 10 dan 90.
- **Dispersion** atau penyebaran data. Untuk keseragaman, keenam jenis pengukuran dispersi dipilih semua.
- **Central Tendency** atau pengukuran pusat data. Untuk keseragaman, klik mouse pada **Mean** dan **Median**.

Di sini pilihan Sum (jumlah usia seluruh karyawan) dan Mode (usia karyawan yang paling sering muncul di variabel usia) dianggap tidak relevan dengan kasus yang ada.

- **Distribution** atau bentuk distribusi data. Untuk keseragaman, klik mouse pada **Skewness** dan **Kurtosis**.

Tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

- Klik mouse pada pilihan **Charts**, maka tampak di layar:



Gambar 2.10 Kotak dialog CHARTS

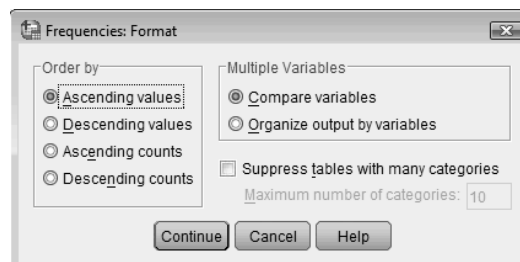
Chart berhubungan dengan jenis grafik yang akan ditampilkan.

Pengisian:

- **Data Type** atau jenis grafik, untuk keseragaman pilih **Bar Charts**.

Tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

Klik mouse pada pilihan **Format**, maka tampak di layar:



Gambar 2.11 Kotak dialog FORMAT

Pilihan ini berhubungan dengan susunan (format) data.

Pengisian:

- **ORDER BY**, pilih **Ascending values** (usia diurutkan dari kecil ke besar). Sedang bagian **MULTIPLE VARIABLES** diabaikan karena di sini hanya dibahas satu variabel saja. Tekan **CONTINUE** untuk kembali ke kotak dialog.

Tekan **OK** jika semua pengisian telah selesai.

Output SPSS dan Analisis

Berikut output dari Frequencies.

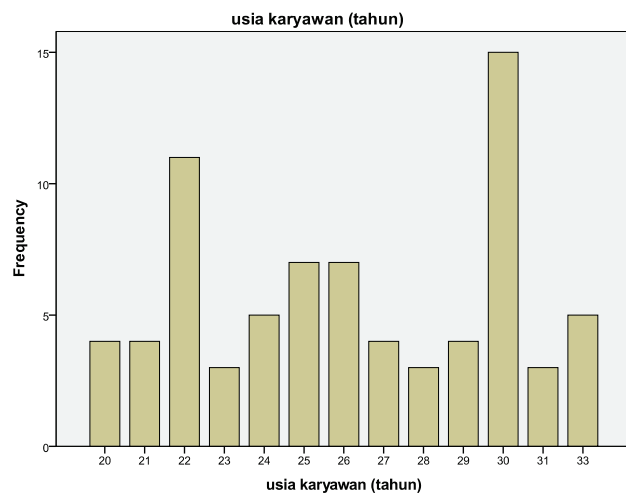
Frequencies

Statistics

usia karyawan (tahun)

N	Valid	75
	Missing	0
Mean		26,24
Std. Error of Mean		,441
Median		26,00
Std. Deviation		3,823
Variance		14,617
Skewness		,052
Std. Error of Skewness		,277
Kurtosis		-1,181
Std. Error of Kurtosis		,548
Range		13
Minimum		20
Maximum		33
Percentiles	10	21,00
	25	22,00
	50	26,00
	75	30,00
	90	31,00

usia karyawan (tahun)					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20	4	5,3	5,3	5,3
	21	4	5,3	5,3	10,7
	22	11	14,7	14,7	25,3
	23	3	4,0	4,0	29,3
	24	5	6,7	6,7	36,0
	25	7	9,3	9,3	45,3
	26	7	9,3	9,3	54,7
	27	4	5,3	5,3	60,0
	28	3	4,0	4,0	64,0
	29	4	5,3	5,3	69,3
	30	15	20,0	20,0	89,3
	31	3	4,0	4,0	93,3
	33	5	6,7	6,7	100,0
	Total	75	100,0	100,0	



Analisis

Karena ada tiga bagian output, maka ada tiga pembahasan.

Output Bagian Pertama (Statistics)

Bagian ini berkenaan dengan deskripsi data untuk variabel Usia karyawan.

- N atau jumlah data yang valid adalah 75 buah, sedangkan data yang hilang (missing) adalah nol. Di sini berarti semua data siap diproses.
- Mean atau rata-rata usia Karyawan adalah 26,24 tahun, dengan standar error adalah 0,44 tahun.

Hal ini bisa ditafsirkan rentang RATA-RATA USIA karyawan terletak pada 3 standar deviasi, atau:

$$\begin{aligned} & 26,24 \text{ tahun} \pm (3 \times 0,44 \text{ tahun}) \\ & = 24,92 \text{ sampai } 27,56 \text{ tahun} \end{aligned}$$

- Median sebesar 26 tahun, menunjukkan bahwa 50% usia karyawan adalah 26 tahun ke atas, dan 50%-nya adalah 26 tahun ke bawah.
- Standar Deviasi adalah 3,82 tahun dan varians -yang merupakan kuadrat dari standar deviasi- adalah 14,62.

Hal ini bisa ditafsirkan rentang USIA karyawan terletak pada 3 standar deviasi, atau:

$$\begin{aligned} & 26,24 \text{ tahun} \pm (3 \times 3,82 \text{ tahun}) \\ & = 14,78 \text{ sampai } 37,7 \text{ tahun} \end{aligned}$$

NB: Bedakan tafsiran RATA-RATA USIA dengan USIA!

- Ukuran Skewness adalah – 0,052. Maka:
Rasio skewness adalah $= \text{nilai skewness} / \text{standar error skewness}$
 $= 0,052 / 0,277 = 0,0675$

Ukuran Kurtosis adalah – 1,181.

Rasio Kurtosis adalah: $\text{nilai Kurtosis} / \text{standar error Kurtosis}$
 $= - 1,181 / 0,548 = - 2,155$ (bisa dianggap mendekati - 2)

Sebagai pedoman, jika rasio kurtosis dan skewness berada di antara – 2 sampai dengan + 2, maka distribusi data adalah normal.

Karena kedua ukuran kemencengan (skewness) dan keruncingan (kurtosis) distribusi usia masih di antara –2 sampai + 2, maka distribusi Usia Karyawan bisa dikatakan Normal.

- Data Minimum adalah 20 tahun, sedangkan data maksimum adalah 33 tahun. Perhatikan bahwa data 20 tahun dan 33 tahun masih *di dalam* batas-batas rentang usia (lihat pembahasan standar deviasi sebelumnya).
- Range adalah data maksimum – data minimum, atau dalam kasus ini:

$$33 - 20 \text{ tahun} = 13 \text{ tahun}$$
- Percentiles atau angka persentil:
 - 10% Karyawan atau sekitar $10\% \times 75 \text{ orang} = 7,5 \text{ orang} \rightarrow 8 \text{ orang}$ mempunyai rata-rata usia di bawah 21 tahun.
 - 25% Karyawan atau sekitar $25\% \times 75 \text{ orang} = 18,75 \text{ orang} \rightarrow 19 \text{ orang}$ mempunyai rata-rata usia di bawah 22 tahun.
 - 50% Karyawan atau sekitar $50\% \times 75 \text{ orang} = 37,5 \text{ orang} \rightarrow 38 \text{ orang}$ mempunyai rata-rata usia di bawah 26 tahun.
Angka 50% persentil sama dengan Median.
 - 75% Karyawan atau sekitar $75\% \times 75 \text{ orang} = 56,25 \text{ orang} \rightarrow 57 \text{ orang}$ mempunyai rata-rata usia di bawah 30 tahun.
 - 90% Karyawan atau sekitar $90\% \times 75 \text{ orang} = 67,5 \text{ orang} \rightarrow 68 \text{ orang}$ mempunyai rata-rata usia di bawah 31 tahun.

Output Bagian Kedua (Tabel Frekuensi USIA)

Penafsiran dilakukan mulai pada baris pertama:

- Baris pertama: Karyawan yang berusia 20 tahun berjumlah (frequency) 4 orang, atau secara persentase dari total jumlah Karyawan adalah $4 / 75 \times 100\% = 5,3\%$.
- Baris kedua: Karyawan yang berusia 21 tahun berjumlah (frequency) 4 orang, atau secara persentase dari total jumlah Karyawan adalah $4 / 75 \times 100\% = 5,3\%$. Jika dihitung secara kumulatif, menjadi: $5,3\%$ (untuk usia 20 tahun) + $5,3\%$ (untuk usia 21 tahun) = $10,7\%$ (ada pembulatan). Hal ini berarti jumlah karyawan berusia 20 dan 21 tahun adalah $10,7\%$ dari total karyawan yang ada.

Demikain seterusnya hingga sampai mencapai 100% kumulatif. Perhatikan hubungan yang erat antara persentase kumulatif dengan persentil.

Output Bagian Ketiga (BAR CHART)

Bagian ketiga menggambarkan grafik data yang telah dibuat frekuensinya, dengan usia karyawan ada pada sumbu X dan frekuensinya ada pada sumbu Y. Terlihat bahwa bentuk bar secara keseluruhan masih bisa dikatakan

berdistribusi normal, walaupun pada usia 22 dan 30 tahun ada data yang *ekstrem*.

Hasil output di atas bisa disimpan dengan nama **frequencies 1**.

KASUS 2

Dari data karyawan di PT MAKMUR RASA akan dibuat Tabel Frekuensi dan melakukan Deskripsi data pada variabel PENGALAMAN KERJA dan TINGKAT PENDIDIKAN KARYAWAN.

Langkah:

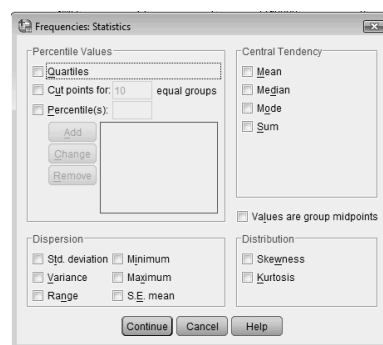
- Buka lembar kerja/ ile **Karyawan**.
- Menu **Analyze → Descriptive Statistics → Frequencies...**



Gambar 2.12 Kotak dialog FREQUENCIES

Pengisian:

- VARIABLE(S), pilih dan masukkan variabel **tingkat_pendidikan** dan **pengalaman_kerja**.
- Klik mouse pada pilihan STATISTICS.



Gambar 2.13 Kotak dialog STATISTICS

Pengisian:

- DISPERSION atau penyebaran data. Untuk keseragaman, pilih **Standard Deviation, Minimum dan Maximum**.
- CENTRAL TENDENCY atau pengukuran pusat data. Untuk keseragaman, klik mouse pada **Mean, Median dan Mode**.

Tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

- Klik mouse pada pilihan CHARTS:
Pilih **Pie Chart** atau grafik ditampilkan berupa lingkaran.
- Klik mouse pada pilihan FORMAT.

Pengisian:

- ORDER BY, pilih **Ascending values**.
- MULTIPLE VARIABLES, pilih **Organize Output by Variables**, yang berarti output akan diurutkan per variabel.

Tekan OK jika semua pengisian telah selesai.

Output SPSS dan Analisis

Berikut output dari Frequencies.

Frequencies

Pendidikan Karyawan

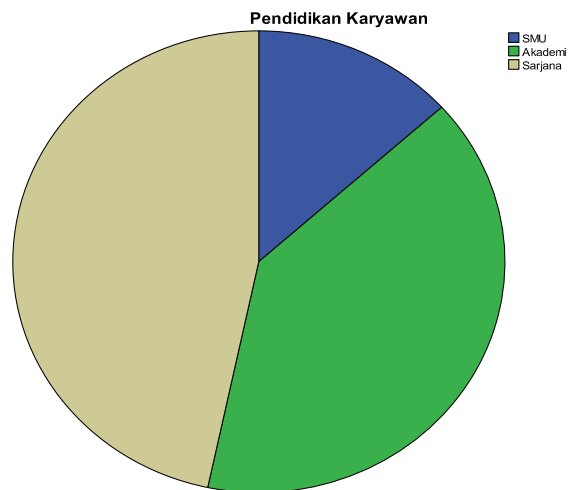
Statistics

Pendidikan Karyawan

N	Valid	75
	Missing	0
Mean		2,3333
Median		2,0000
Mode		3,00
Std. Deviation		,70391
Minimum		1,00
Maximum		3,00

Pendidikan Karyawan

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SMU	10	13,3	13,3	13,3
	Akademi	30	40,0	40,0	53,3
	Sarjana	35	46,7	46,7	100,0
	Total	75	100,0	100,0	



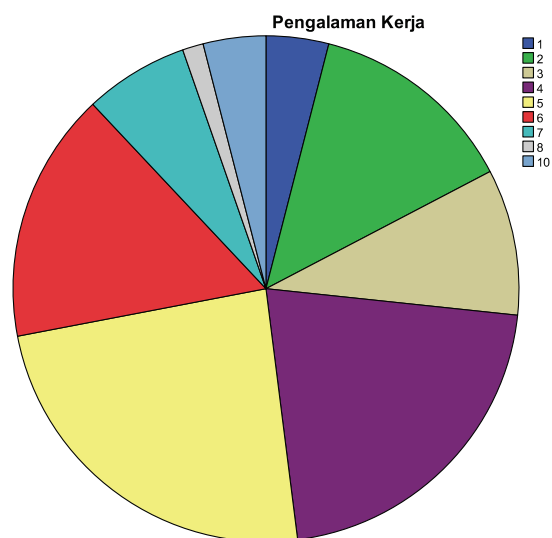
Pengalaman Kerja

Statistics

Pengalaman Kerja

N	Valid	75
	Missing	0
Mean		4,57
Median		5,00
Mode		5
Std. Deviation		1,960
Minimum		1
Maximum		10

Pengalaman Kerja					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	4,0	4,0	4,0
	2	10	13,3	13,3	17,3
	3	7	9,3	9,3	26,7
	4	16	21,3	21,3	48,0
	5	18	24,0	24,0	72,0
	6	12	16,0	16,0	88,0
	7	5	6,7	6,7	94,7
	8	1	1,3	1,3	96,0
	10	3	4,0	4,0	100,0
Total		75	100,0	100,0	



Analisis

Terlihat output menjelaskan per variabel.

Output Bagian Pertama (Variabel PENDIDIKAN):

Karena variabel PENDIDIKAN KARYAWAN adalah data kualitatif, maka sesungguhnya yang *relevan* untuk deskripsi data berdasar output bagian pertama di atas adalah:

- Mode/Modus, yaitu jenis Pendidikan yang paling banyak adalah 3 (Sarjana).
- Minimum dan Maksimum, yang berarti jenis pendidikan yang paling sedikit (1 atau SMU) dan terbanyak (3 atau Sarjana). Hal ini terbukti dari tabel output kedua, yang menunjukkan jumlah Karyawan berpendidikan SMU hanya 10 orang, sedang Sarjana sebanyak 35 orang.

Sedangkan statistik lain, seperti rata-rata pendidikan (Mean) yang sebesar 2,33 adalah tidak relevan, karena 2,333 tidak mengacu pada kode apa pun. Demikian juga dengan standar deviasi.

Output Bagian Ketiga (Variabel PENGALAMAN KERJA):

Karena variabel ini termasuk data kuantitatif, maka bisa digunakan statistik seperti pembahasan variabel Usia terdahulu.

- N atau jumlah data yang valid adalah 75 buah.
- Mean atau rata-rata Pengalaman Kerja Karyawan adalah 4,57 tahun, dengan standar deviasi adalah 1,96 tahun.
- Median sebesar 5 tahun, menunjukkan bahwa sekitar 50% Pengalaman kerja karyawan adalah 5 tahun ke atas, dan 50%-nya adalah 5 tahun ke bawah.
- Mode adalah 5 tahun, atau data 5 paling sering keluar pada perhitungan data.

Output Bagian Keempat (Tabel Frekuensi PENGALAMAN KERJA):

Penafsiran dilakukan mulai pada baris pertama:

- Baris pertama: Karyawan dengan pengalaman kerja 1 tahun berjumlah 3 orang, atau $3 / 75 * 100\% = 4\%$.
- Baris kedua: Karyawan dengan pengalaman kerja 2 tahun berjumlah 10 orang, atau $10 / 75 * 100\% = 13,33\%$. Jika dihitung secara kumulatif, menjadi 17,3%. Demikian seterusnya hingga sampai mencapai 100% kumulatif.

Terlihat juga dua bagan bentuk Pie (lingkaran) yang menggambarkan komposisi Pendidikan dan Pengalaman Kerja karyawan.

Hasil output di atas bisa disimpan dengan nama **frequencies 2**.

2.5 DESCRIPTIVE

Fasilitas DESCRIPTIVE mempunyai kesamaan dengan fasilitas untuk statistik deskriptif lainnya pada SPSS, seperti FREQUENCIES atau CASE SUMMARIES. Perbedaan utama hanya pada kemampuan menghasilkan skor standardisasi nilai z. Nilai z digunakan khususnya untuk menstandarisasi sejumlah nilai yang mempunyai perbedaan angka besar.

Data yang digunakan pada DESCRIPTIVE adalah data kuantitatif, walaupun data kualitatif dapat juga menjadi input untuk proses pengolahan data.

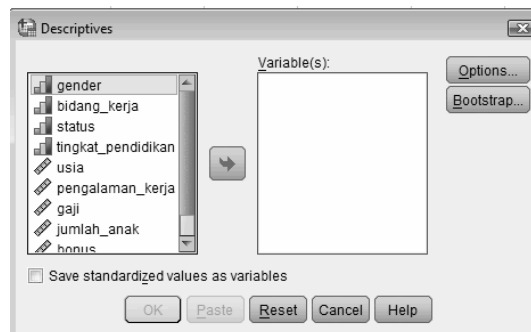
KASUS 1

Dari data karyawan di PT MAKMUR RASA, deskripsikan variabel GAJI yang ada pada data tersebut.

Langkah-langkah:

- Buka file **Karyawan**.
- Menu **Analyze → Descriptive Statistics → Descriptives....**

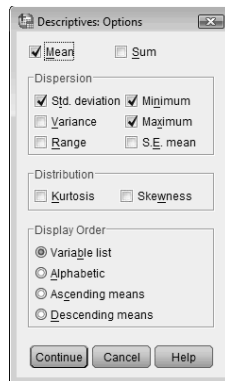
Tampak di layar:



Gambar 2.14 Kotak dialog DESCRIPTIVE

Pengisian:

- VARIABLE(S) atau variabel yang akan dimasukkan. Pilih variabel **gaji**.
- OPTIONS atau pilihan untuk pengerjaan deskripsi data. Klik pilihan tersebut, tampak di layar:



Gambar 2.15 Kotak dialog OPTIONS

Pengisian, untuk keseragaman pilih:

- **Mean, Standard Deviation, Minimum dan Maximum** (sudah sebagai default).
- **Sum.**

Kemudian tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

Abaikan bagian yang lain dan tekan OK jika semua pengisian telah selesai.

Output SPSS dan Analisis

Berikut output dari Deskriptif.

Descriptives

Descriptive Statistics						
	N	Minimum	Maximum	Sum	Mean	Std. Deviation
Gaji per bulan	75	930500	1830000	98346000	1311280,00	191126,430
Valid N (listwise)	75					

Analisis

Tujuh puluh lima karyawan PT MAKMUR RASA mempunyai gaji rata-rata Rp. 1.311.280,- dengan standar deviasi Rp. 191.126,43. Total Gaji

seluruhnya adalah Rp. 98.346.000,- dengan gaji minimum Rp. 930.500,- dan maksimum Rp. 1.830.000,-. Standar deviasi yang cukup besar (lebih dari 14,5% dari Mean) menunjukkan adanya variasi yang besar, atau adanya kesenjangan yang cukup besar dari gaji terendah dan tertinggi.

Simpan output tersebut dengan nama **DESCRIPTIVE 1**.

KASUS 2

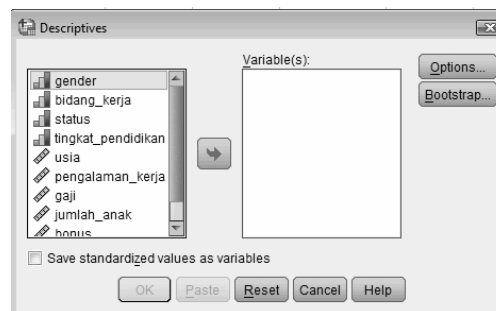
Deskripsikan variabel **USIA** pada data karyawan, dan apakah pada variabel usia tersebut terdapat data yang ‘ekstrem’?

Pengertian ekstrem di sini adalah adanya data yang jauh di atas atau di bawah rata-ratanya (outlier).

Langkah:

- Buka file **Karyawan**.
- Menu **Analyze → Descriptive Statistics → Descriptives....**

Tampak di layar:



Gambar 2.16 Kotak dialog DESCRIPTIVE

Pengisian:

- Tekan tombol **Reset** untuk menghapus semua input yang sebelumnya ada (jika di layar terdapat variabel yang sudah terinput ke kotak dialog).
- VARIABLE(S). Masukkan variabel **usia**.
- Aktifkan pilihan **Save standardized values as variables** untuk mengetahui penyimpangan data. Dengan mengaktifkan pilihan ini, otomatis ada sebuah variabel baru di data Karyawan.
- OPTIONS. Untuk keseragaman pilih **Mean, Standard Deviation, Minimum, dan Maximum**.

Kemudian tekan **Continue**. Abaikan bagian yang lain dan tekan OK jika semua pengisian telah selesai.

Output SPSS dan Analisis

Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
usia karyawan (tahun)	75	20	33	26,24	3,823
Valid N (listwise)	75				

Analisis

Output Deskriptif

Tujuh puluh lima karyawan mempunyai usia rata-rata 26,24 tahun dengan standar deviasi 3,82 tahun.

Output mencari adanya outlier

Jika dilihat pada Data Editor SPSS (file Karyawan), sekarang muncul variabel baru, yaitu **zusia** sebagai berikut (hanya ditampilkan variabel Usia dan zusia).

	usia	zusia
1	23	- 8,4744
2	21	- 1,37056
3	21	- 1,37056
4	22	- 1,10900
5	24	- ,58589

.....dan seterusnya sampai 75 data.

Keterangan:

Jika suatu data berdistribusi normal, maka suatu nilai bisa distandardisasi dengan nilai z:

$$z = \frac{x_i - \bar{x}}{s}$$

Di mana:

X_i = nilai data ke-i (data 1, 2, 3, ... sampai 75)

X = Mean data (dalam kasus adalah 26,24 tahun)

S = standar deviasi (dalam kasus adalah 3,82 tahun)

Sebagai contoh, lihat pada baris pertama, yaitu usia 23 tahun, maka nilai z-nya adalah:

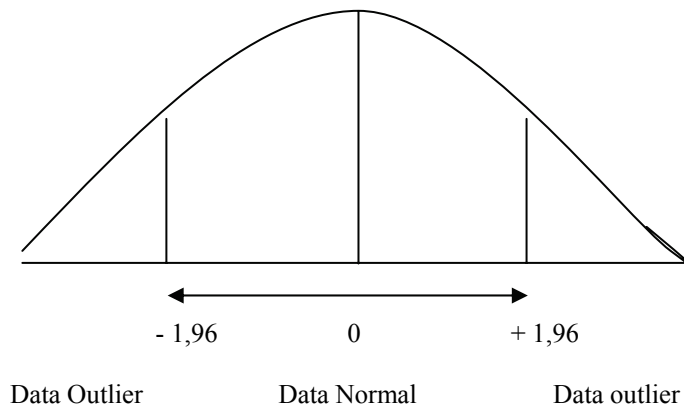
$$z = \frac{23 - 26,24}{3,82} = -8,4744 \text{ (sama dengan output SPSS)}$$

Demikian untuk data yang lain.

Melihat Data yang Menyimpang (Outlier)

Untuk data normal, nilai z akan terletak (pada taraf signifikansi 5% atau luas kurva 0,475) antara $-1,96$ sampai $+1,96$.

Gambar:



Data outlier akan mempunyai nilai lebih besar dari $+1,96$ atau lebih kecil dari $-1,96$. Dari nilai variabel usia terlihat TIDAK ADA data yang termasuk data outlier, atau bisa dikatakan semua data variabel Usia adalah normal.

Simpan output dengan nama **DESCRIPTIVE 2**; simpan pula file karyawan yang sekarang telah mempunyai satu variabel baru berisi nilai z dengan nama **karyawan z value**.

2.6 EXPLORE

Menu EXPLORE mempunyai fungsi sama dengan menu DESCRIPTIVE sebelumnya, yaitu melakukan deskripsi data dan menguji apakah ada data yang outlier. Hanya di sini pengujian kenormalan data dilakukan lewat:

- tampilan diagram, seperti Boxplot dan Normal Probability Plot.
- uji tertentu, seperti uji Shapiro Wilks dan Lilliefors.

Data yang digunakan untuk menu ini adalah data kualitatif dan data kuantitatif, yang digunakan secara bersamaan.

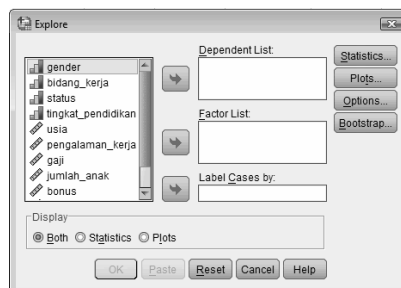
KASUS 1

Dari data karyawan PT MAKMUR RASA, akan dilakukan eksplorasi data pada variabel BIDANG KERJA dengan PENGALAMAN KERJA karyawan.

Perhatikan bahwa variabel bidang_kerja adalah variabel kualitatif (karena ada kode), sedangkan variabel pengalaman_kerja adalah data kuantitatif.

Langkah:

- Buka file **Karyawan**.
- Menu **Analyze → Descriptive Statistics → Explore....**

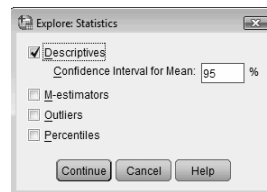


Gambar 2.17 Kotak dialog **EXPLORE**

Pengisian:

- **DEPENDENT LIST** atau variabel tergantung. Pilih variabel **pengalaman_kerja**.

- FACTOR LIST atau faktor yang akan dimasukkan. Pilih variabel dengan tipe kualitatif, atau dalam soal ini adalah **bidang_kerja**.
- Abaikan pengisian LIST CASES BY.
- Klik pada tombol STATISTICS, tampak di layar:



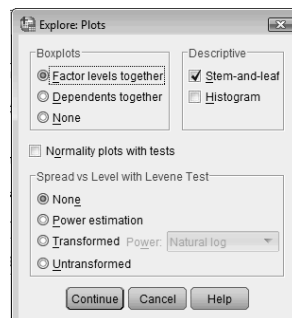
Gambar 2.18 Kotak dialog STATISTICS

Pengisian:

Pilih (aktifkan) **Descriptives**, **M-estimators** dan **Outliers**.

Kemudian tekan **Continue** untuk melanjutkan proses berikutnya.

- Klik pada pilihan PLOTS, tampak di layar:



Gambar 2.19 Kotak dialog PLOTS

Pengisian:

- Untuk keseragaman, pilihan diisi sesuai default, yaitu pada Boxplot adalah **factor levels together** dan pada descriptive adalah **stem and leaf**.

Tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

- Pada bagian DISPLAY, pilih **Both** yang berarti baik statistics maupun Plots akan digunakan.

Tekan OK jika semua pengisian telah selesai.

Explore
Bidang

Case Processing Summary

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Pengalaman Kerja	Administrasi	15	100,0%	0	,0%	15	100,0%
	Personalia	8	100,0%	0	,0%	8	100,0%
	Produksi	12	100,0%	0	,0%	12	100,0%
	Marketing	27	100,0%	0	,0%	27	100,0%
	Keuangan	13	100,0%	0	,0%	13	100,0%

Descriptives

Bidang			Statistic	Std. Error
Pengalaman Kerja	Administrasi	Mean	4,47	,435
		95% Confidence Interval for Mean	3,53	
		Lower Bound		
		Upper Bound	5,40	
		5% Trimmed Mean	4,46	
		Median	4,00	
		Variance	2,838	
		Std. Deviation	1,685	
		Minimum	2	
		Maximum	7	
		Range	5	

		Interquartile Range	3	
		Skewness	,273	,580
		Kurtosis	-,866	1,121
Personalia	Mean		6,13	,639
	95% Confidence Interval for Mean	Lower Bound	4,61	
		Upper Bound	7,64	
	5% Trimmed Mean		6,03	
	Median		6,00	
	Variance		3,268	
	Std. Deviation		1,808	
	Minimum		4	
	Maximum		10	
	Range		6	
	Interquartile Range		2	
	Skewness		1,496	,752
	Kurtosis		3,115	1,481
Produksi	Mean		4,00	,508
	95% Confidence Interval for Mean	Lower Bound	2,88	
		Upper Bound	5,12	
	5% Trimmed Mean		4,00	
	Median		4,00	
	Variance		3,091	
	Std. Deviation		1,758	
	Minimum		1	

		Maximum	7	
		Range	6	
		Interquartile Range	3	
		Skewness	,120	,637
		Kurtosis	-,504	1,232
Marketing	Mean		4,78	,404
	95% Confidence Interval for Mean	Lower Bound	3,95	
		Upper Bound	5,61	
	5% Trimmed Mean		4,64	
	Median		5,00	
	Variance		4,410	
	Std. Deviation		2,100	
	Minimum		2	
	Maximum		10	
	Range		8	
	Interquartile Range		1	
	Skewness		,936	,448
	Kurtosis		1,383	,872
Keuangan	Mean		3,85	,517
	95% Confidence Interval for Mean	Lower Bound	2,72	
		Upper Bound	4,97	
	5% Trimmed Mean		3,88	
	Median		4,00	
	Variance		3,474	

Std. Deviation	1,864	
Minimum	1	
Maximum	6	
Range	5	
Interquartile Range	4	
Skewness	-,376	,616
Kurtosis	-1,359	1,191

M-Estimators

Bidang		Huber's M- Estimator ^a	Tukey's Biweight ^b	Hampel's M- Estimator ^c	Andrews' Wave ^d
Pengalaman Kerja	Administrasi	4,29	4,27	4,37	4,27
	Personalia	5,84	5,60	5,79	5,59
	Produksi	3,91	3,91	3,96	3,91
	Marketing	4,67	4,45	4,54	4,44
	Keuangan	3,88	3,90	3,85	3,90

a. The weighting constant is 1,339.

b. The weighting constant is 4,685.

c. The weighting constants are 1,700, 3,400, and 8,500

d. The weighting constant is $1,340 \cdot \pi$.

Extreme Values^f

Bidang				Case Number	Value
Pengalaman Kerja	Administrasi	Highest	1	9	7
			2	10	7
			3	14	7
			4	15	6
			5	4	5 ^a
		Lowest	1	2	2
			2	1	2
			3	13	3
			4	3	3
			5	12	4 ^b
	Personalia	Highest	1	22	10
			2	21	7
			3	17	6
			4	18	6 ^c
		Lowest	1	16	4
			2	23	5
			3	19	5
			4	20	6 ^d
	Produksi	Highest	1	35	7
			2	30	6
			3	33	6
			4	26	5
			5	29	4 ^e
		Lowest	1	25	1

		2	27	2
		3	31	3
		4	28	3
		5	24	3
Marketing	Highest	1	48	10
		2	50	10
		3	61	8
		4	45	6
		5	51	6 ^c
	Lowest	1	62	2
		2	54	2
		3	49	2
		4	40	2
		5	36	2
Keuangan	Highest	1	64	6
		2	69	6
		3	70	6
		4	67	5
		5	68	5 ^a
	Lowest	1	71	1
		2	65	1
		3	66	2
		4	63	2
		5	72	3

- a. Only a partial list of cases with the value 5 are shown in the table of upper extremes.
- b. Only a partial list of cases with the value 4 are shown in the table of lower extremes.
- c. Only a partial list of cases with the value 6 are shown in the table of upper extremes.
- d. Only a partial list of cases with the value 6 are shown in the table of lower extremes.
- e. Only a partial list of cases with the value 4 are shown in the table of upper extremes.
- f. The requested number of extreme values exceeds the number of data points. A smaller number of extremes is displayed.

Pengalaman Kerja Stem-and-Leaf Plots

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Administrasi

Frequency Stem & Leaf

2,00	2 . 00
2,00	3 . 00
5,00	4 . 00000
2,00	5 . 00
1,00	6 . 0
3,00	7 . 000

Stem width: 1
Each leaf: 1 case(s)

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Personalia

Frequency Stem & Leaf

1,00	4 . 0
2,00	5 . 00
3,00	6 . 000
1,00	7 . 0
1,00	Extremes (>=10,0)

Stem width: 1

Each leaf: 1 case(s)

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Produksi

Frequency Stem & Leaf

1,00	0 . 1
4,00	0 . 2333
4,00	0 . 4445
3,00	0 . 667

Stem width: 10

Each leaf: 1 case(s)

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Keuangan

Frequency Stem & Leaf

2,00	1 . 00
2,00	2 . 00
1,00	3 . 0
2,00	4 . 00
3,00	5 . 000
3,00	6 . 000

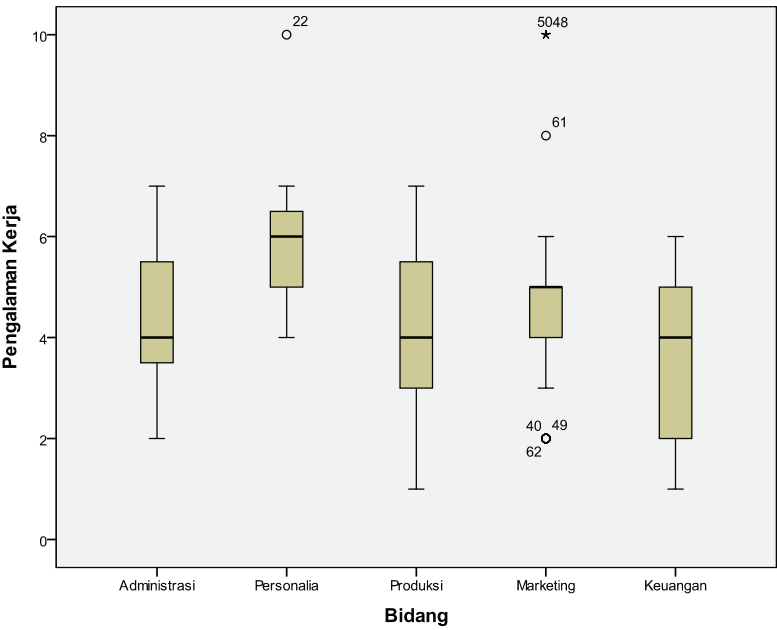
Stem width: 1

Each leaf: 1 case(s)

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Marketing

Frequency Stem & Leaf

5,00 Extremes (= <2,0)
1,00 3 . 0
 ,00 3 .
5,00 4 . 00000
 ,00 4 .
10,00 5 . 0000000000
 ,00 5 .
3,00 6 . 000
3,00 Extremes (>=8,0)
Stem width: 1
Each leaf: 1 case(s)



Analisis

Output Case Processing Summary

Semua data Karyawan (75 orang) valid (100%).

Output Descriptives

Bagian ini mengenai ringkasan statistik deskripsi dari Gaji karyawan berdasar bidang pekerjaannya. Penjelasan lengkap bisa mengacu pada pembahasan Frequencies dan Descriptive di depan.

Output M-Estimators

M-Estimator (Mean Estimator) bisa sebagai alternatif pengukuran pusat, yaitu dengan memberi bobot (weight) pada data. Sebagai contoh, menurut pengukuran Huber, rata-rata Pengalaman Kerja berdasar bidang adalah:

- Administrasi: 4,3 tahun
- Personalia: 5,84 tahun
- Produksi: 3,91 tahun
- Marketing: 4,67 tahun
- Keuangan: 3,88 tahun

Demikian seterusnya untuk jenis pengukuran yang lain (Tukey, Hampel, Andrew). Terlihat pengukuran Mean oleh Humpel memberi hasil yang relatif lebih besar dibanding yang lain.

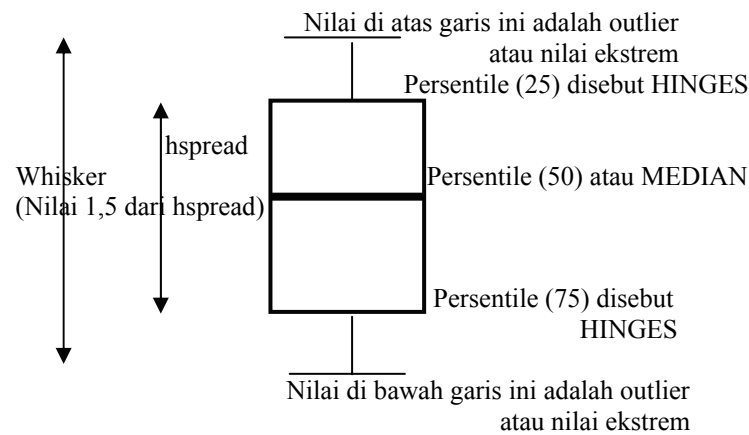
Output Extreme Value:

Output ini menampilkan 5 nilai tertinggi dan 5 nilai terendah dari 75 data karyawan berdasar Bidang kerja. Sebagai contoh, untuk bidang administrasi, nilai terkecil adalah 2 dan 3 tahun, dengan nomor kasus adalah 2,1,13,3 dan 12. Sedang untuk nilai terbesar adalah 6 dan 7 tahun, dengan nomor kasus adalah 9,10,14,15 dan 4. Demikian seterusnya untuk bidang kerja yang lain.

Output BOXPLOT:

Boxplot adalah kotak pada gambar berwarna merah (atau mungkin warna yang lain) dengan garis tebal horizontal di kotak tersebut. Kotak merah tersebut memuat 50% data, atau mempunyai batas persentil ke-25 dan ke-75 (lihat pembahasan interquartile mean). Sedangkan garis tebal hitam adalah Median data.

Berikut gambar Boxplot teoritis.



Keterangan:

- nilai lebih dari 1,5 hspread (tinggi Boxplot) ditandai dengan ‘o’ dan disebut outlier.
- nilai lebih dari 3 hspread (tinggi Boxplot) ditandai dengan ‘*’ dan disebut extrem value atau ‘far outside value’.
- Jika garis hitam atau tanda Median terletak persis di tengah Boxplot, maka distribusi data adalah normal. Jika berada di sebelah atas, distribusi menceng ke kiri. Jika di sebelah bawah, distribusi menceng ke kanan.

Analisis

- Rata-rata pengalaman kerja karyawan bidang Personalia adalah paling tinggi, kemudian disusul oleh bidang Marketing, kemudian ketiga bidang sisanya. Hal ini ditunjukkan oleh letak bidang Personalia yang paling tinggi dibanding bidang lainnya.
- Dari posisi garis hitam (Mean) dari masing-masing bidang, terlihat bahwa data bidang marketing sangat menceng ke kiri (letak garis hitam di atas), sedang bidang administrasi menceng ke kanan (letak garis hitam di bawah).
- Ada 9 nilai outlier dan ekstrem, yaitu satu di bidang personalia (kasus/ data nomor 22), dan delapan pada Marketing (data nomor 50, 48, 40, 49, 61 dan 62). Jika pada Boxplot nomor tidak jelas, bisa klik ganda mouse, maka akan tampak Chart Editor, dan kemudian gambar bisa diperbesar sesuai keperluan.

Namun, jika belum terlihat jelas, bisa dilihat pada pembahasan stem and leaf atau bagian *Extreme Values*.

Ouput STEM AND LEAF:

Contoh stem and leaf Personalia.

Pengalaman Kerja Stem-and-Leaf Plot for
bidang_kerja= Personalia

Frequency	Stem &	Leaf
1,00	4 .	0
2,00	5 .	00
3,00	6 .	000
1,00	7 .	0
1,00	Extremes	(>=10,0)

Stem width: 1
Each leaf: 1 case(s)

Analisis:

- Pada baris pertama, ada satu karyawan yang mempunyai pengalaman kerja 4 tahun (stem = 4). Leaf atau cabang yang bernilai 0, berarti hanya seorang saja, dan pengalaman kerja tidak ada desimal (0).
- Pada baris kedua, ada dua karyawan yang mempunyai pengalaman kerja 5 tahun (stem = 5). Leaf atau cabang yang bernilai 00, berarti ada dua orang, dan pengalaman kerja tidak ada desimal (0).

Demikian seterusnya untuk data yang lain.

Untuk data dengan tulisan 'EXTREMES' menunjukkan ada satu karyawan di bagian personalia dengan pengalaman kerja sama atau di atas (≥ 10 tahun). Jika dilihat di BOXPLOT, karyawan tersebut mempunyai nomor baris (case) adalah 22, yang pada data karyawan mempunyai pengalaman kerja tepat 10 tahun.

Simpan output dengan nama **EXPLORE 1**.

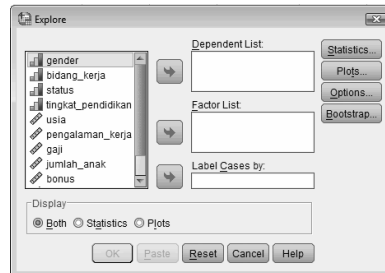
KASUS 2

Dari data karyawan PT MAKMUR RASA akan dilakukan eksplorasi data pada variabel STATUS dengan USIA karyawan.

Langkah:

- Buka file **Karyawan**.

- Menu **Analyze** → **Descriptive Statistics** → **Explore....**



Gambar 2.20 Kotak dialog EXPLORE

Pengisian:

- **DEPENDENT LIST.** Pilih variabel **usia**.
- **FACTOR LIST,** pilih **status**.
- Abaikan pengisian **LIST CASES BY**.
- Pada bagian **DISPLAY**, pilih **Plot** yang berarti hanya Plots yang akan digunakan.
- Klik pada pilihan **PLOTS**, tampak di layar kotak dialog **PLOTS**.
Untuk keseragaman, aktifkan pilihan **Normality Plot with tests**, dan **NON AKTIFKAN** bagian **Stem and Leaf** bila masih aktif (ada tanda ✓).
Abaikan bagian yang lain. Kemudian tekan **Continue** setelah selesai input untuk melanjutkan proses berikutnya.

Tekan OK jika semua pengisian telah selesai.

Output:

Explore

Status Karyawan

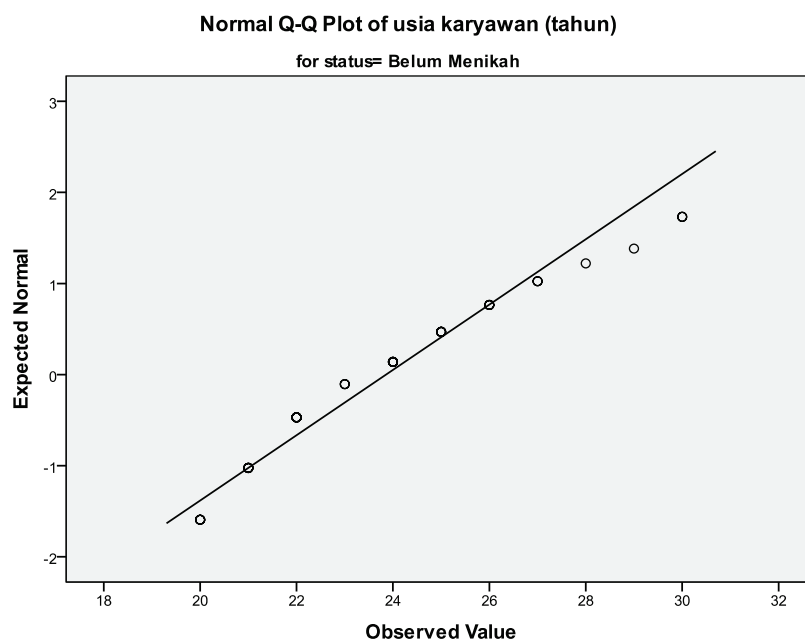
Tests of Normality							
Status Karyawan		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
usia karyawan (tahun)	Belum Menikah	,176	35	,008	,933	35	,033
	Menikah						
	Menikah	,215	40	,000	,913	40	,005

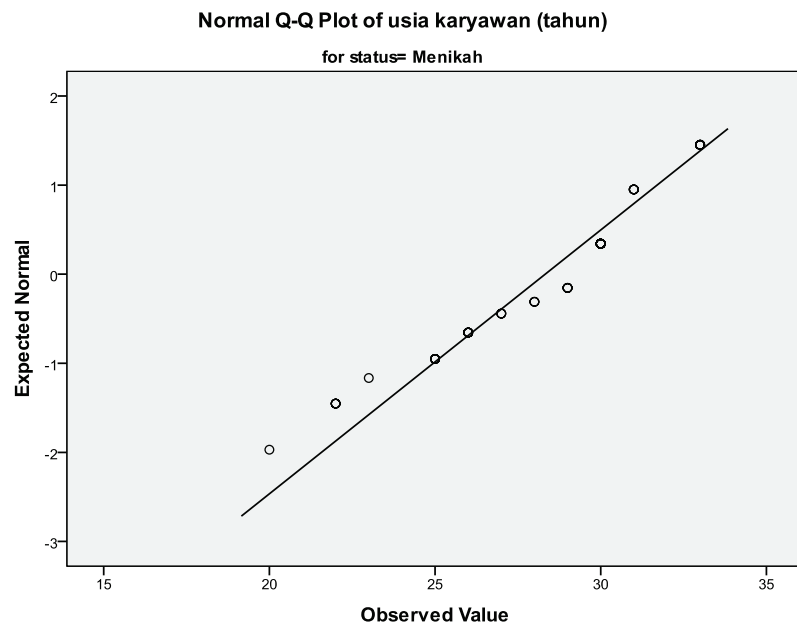
Tests of Normality

Status Karyawan		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
usia karyawan (tahun)	Belum	,176	35	,008	,933	35	,033
	Menikah						
	Menikah	,215	40	,000	,913	40	,005

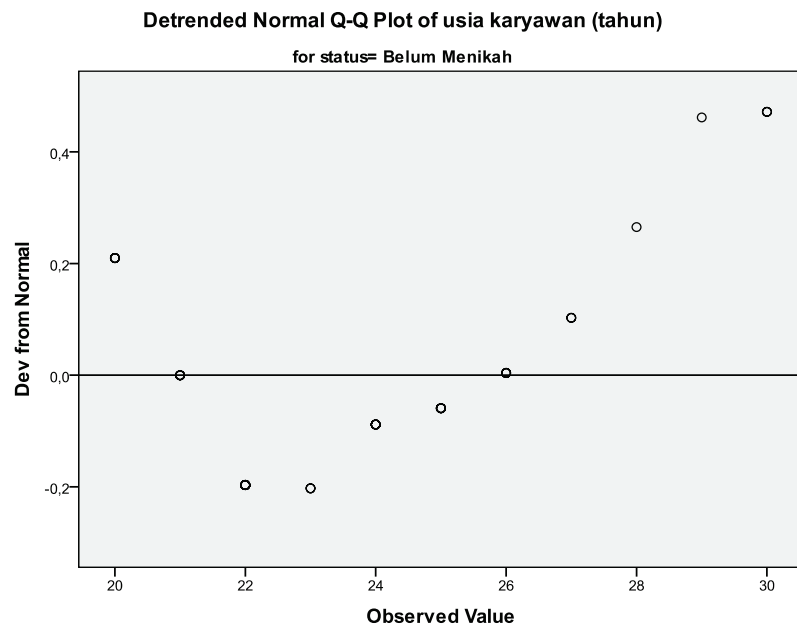
a. Lilliefors Significance Correction

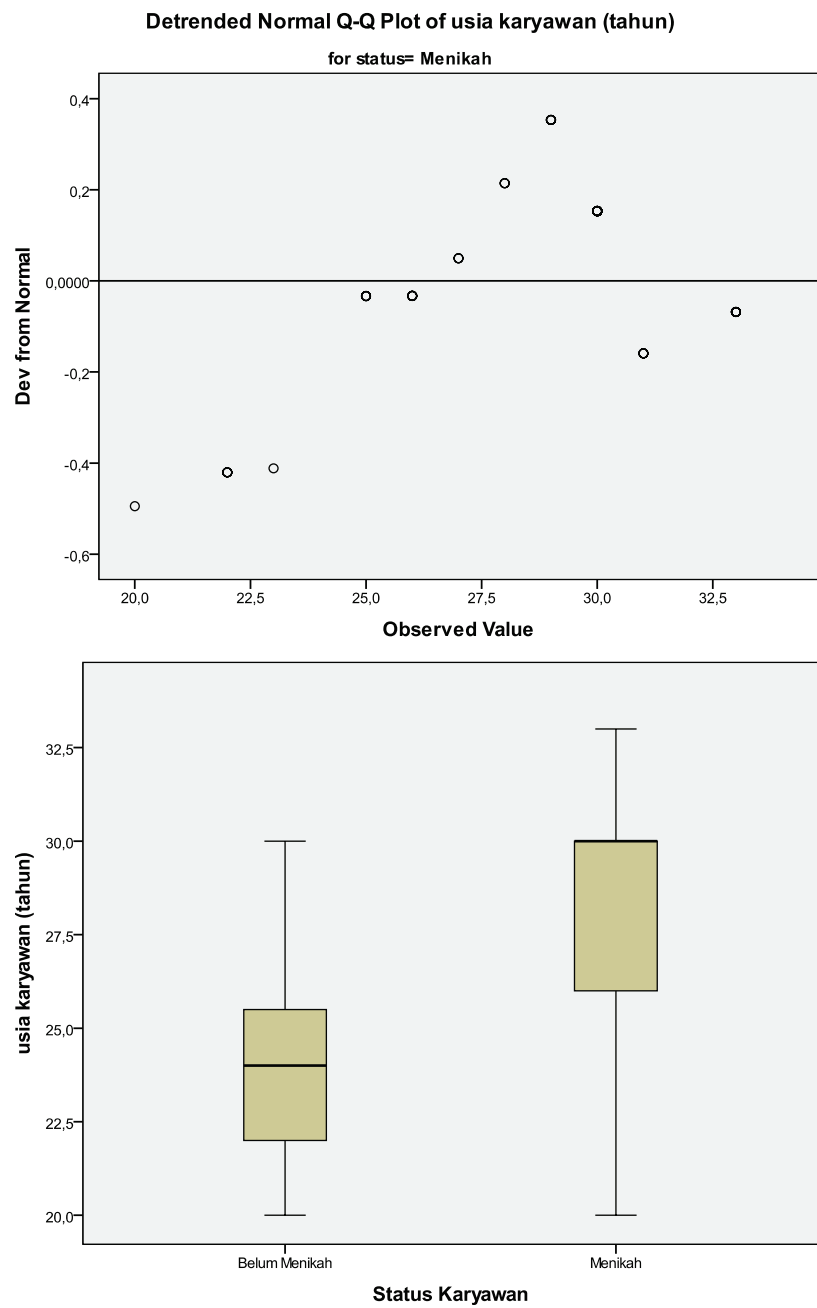
USIA Normal Q-Q Plots





Detrended Normal Q-Q Plots





Analisis

Output Case Processing Summary:

Semua data Karyawan (75 orang) valid (100%).

Output Test of Normality:

Bagian ini akan menguji normal tidaknya sebuah distribusi data.

Pedoman pengambilan keputusan:

- Nilai Sig. atau signifikansi atau nilai probabilitas $< 0,05$, Distribusi adalah tidak normal.
- Nilai Sig. atau signifikansi atau nilai probabilitas $> 0,05$, Distribusi adalah normal.

Analisis

Pada hasil uji Kolmogorov Smirnov dan Shapiro Will didapatkan untuk Karyawan yang berstatus Menikah atau berstatus belum menikah, distribusi Usia karyawan adalah Tidak Normal. Hal ini bisa dilihat pada tingkat signifikansi kedua alat uji, yang jauh di bawah 0,05 (5%).

Output untuk menguji normalitas dengan plot (Q-Q plot):

Jika suatu distribusi data normal, maka **data akan tersebar di sekeliling garis**. Terlihat bahwa baik untuk status menikah atau belum menikah, pola data tersebar cukup jauh melewati garis standar, yang berarti data tidak bisa dikatakan berdistribusi normal.

Output untuk menguji normalitas dengan plot (Detrended Normal Q-Q plot):

Output ini untuk mendeteksi pola-pola dari titik-titik yang bukan bagian dari kurva normal. Terlihat bahwa baik pada status menikah atau belum menikah, sebagian besar data terpola jauh di garis tengah yang ada. Hal ini membuktikan bahwa distribusi data adalah tidak normal.

Simpan output dengan nama **EXPLORE 2**.

